SOP-GPU: Influence of Solvent-Induced Hydrodynamic Interactions on Dynamic Structural Transitions in Protein Assemblies

Andrey Alekseenko,[a,b] Olga Kononova,[a,b] Yaroslav Kholodov,[c,d] Kenneth A. Marx,[a] and Valeri Barsegov*[a,b]

Hydrodynamic interactions (HI) are incorporated into Langevin dynamics of the C_α-based protein model using the Truncated Expansion approximation (TEA) to the Rotne–Prager–Yamakawa diffusion tensor. Computational performance of the obtained GPU realization demonstrates the model's capability for describing protein systems of varying complexity (10^2–10^5 residues), including biological particles (filaments, virus shells). Comparison of numerical accuracy of the TEA versus exact description of HI reveals similar results for the kinetics and thermodynamics of protein unfolding. The HI speed up and couple biomolecular transitions through cross-communication among protein domains, which result in more collective displacements of structure elements governed by more deterministic (less variable) dynamics. The force-extension/deformation spectra from nanomanipulations *in silico* exhibit sharper force signals that match well the experimental profiles. Hence, biomolecular simulations without HI overestimate the role of tension/stress fluctuations. Our findings establish the importance of incorporating implicit water-mediated many-body effects into theoretical modeling of dynamic processes involving biomolecules. © 2016 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

Introduction

The solvent environment affects the dynamic properties of biomolecules, but the issue of hydrodynamic coupling between various segments within the same biological macromolecule has not received much attention. Although hydrodynamic interactions (HI) affect the properties of condensed phase systems, there are only a handful of reports in the literature on the effects of these interactions in proteins. Recent theoretical studies have mainly focused on how hydrodynamic effects influence the diffusion of biomolecules,[1,2] protein ensembles' association,[3,4] and protein folding.[5] Very few studies have focused on how hydrodynamic interactions affect the kinetics and thermodynamics of protein unfolding. In a recent study,[6] the investigators addressed the role of hydrodynamic interactions on protein unfolding, but the example chosen was a small protein-ubiquitin. Therefore, a number of interesting questions remain. For example, it is not known how hydrodynamic interactions influence the unfolding pathways and mechanism(s) in multidomain proteins under tension. Little is known about the extent of domain–domain coupling in multidomain proteins in the presence of hydrodynamic interactions. More intriguing questions concern the effect of hydrodynamic interactions on the biophysical properties of large protein complexes. It is not yet known how hydrodynamic interactions modulate any large system's dynamic response to external mechanical factors, where examples of such important biological assemblies include microtubules, cell organelles, and animal and plant virus particles, etc.

Langevin dynamics simulations based on simplified models of biomolecules are widely used to access large-amplitude and long-timescale molecular motions. In these simulations, the description of solvent effects is typically limited to stochastic collisions implicitly reflecting the thermal environment. These are modeled by random forces acting on each particle and in each dimension x, y, and z in the three-dimensional space. Although the collective behavior of solvent molecules can dramatically alter the dynamics of the whole system, in coarse-grained modeling approaches this behavior is usually neglected. Here, we employ the Langevin dynamics simulations of the C_α-based Self Organized Polymer (SOP) model of a polypeptide chain[7,8] to explore the influence of solvent

[a] A. Alekseenko, O. Kononova, K. A. Marx, V. Barsegov
Department of Chemistry, One University ave, University of Massachusetts, Lowell, Massachusetts 01854
E-mail: Valeri_Barsegov@uml.edu

[b] A. Alekseenko, O. Kononova, V. Barsegov
9 Institutskiy per, Department of Control and Applied Mathematics Moscow Institute of Physics and Technology, Moscow region 141700, Russia

[c] Y. Kholodov
1 Universitetskaya st, Department of Information Systems Innopolis University, Innopolis, Tatarstan 420500, Russia

[d] Y. Kholodov
19/18, 2-nd Brestskaya st, Department of Applied Mathematics Institute of Computer Aided Design of the Russian Academy of Sciences, Moscow 123056, Russia

induced many-body effects on dynamic structural transitions in protein assemblies. A variety of biochemical systems have been described using this approach.[7,9–11]

We have ported the SOP-based simulations to Graphics Processing Units (GPUs).[12,13] The computational acceleration attained on a GPU with the SOP-GPU program (see Fig. S1 in the Supporting Information) has enabled us to address the biological N-body problem and to probe dynamic structural transitions in: fibrin monomer and oligomers ($N \approx 10^3$–10^4 amino acids),[14,15] microtubule polymers ($N \approx 10^5$),[16] and protein shells of the virus particle CCMV and the bacteriophage HK97 ($N \approx 10^4$–10^5).[13,17,18] Here, we employ the SOP-model to account for the hydrodynamic interactions between the C_α-particles with all the computational subroutines fully implemented on a GPU. In the past decade, GPUs have transformed from being an obscure rapid visualization oriented technical development to becoming integral components of modern supercomputers with wide software support. The utilization of GPUs for Molecular Dynamics (MD) based scientific applications began in 2007,[19] but now successful MD simulation packages have partial or full GPU support including HOOMD-blue,[20] ACEMD,[21] AMBER,[22,23] and GROMACS.[24]

A simple method to account for hydrodynamic interactions in Langevin dynamics simulations was proposed by Ermak and McCammon.[25] A downside of this approach is the requirement to perform computationally expensive Cholesky decomposition to obtain the correlation matrix for random forces at each step of numerical integration. Several methods have been introduced to leverage this difficulty, including the Fixman approximation[26] and Krylov subspace methods,[27] but the issues remain. Here, we propose to use an approximate treatment of the hydrodynamic effects based on the Truncated Expansion approximation (TEA).[28] The obtained realization of the SOP model with hydrodynamics is fully implemented on a GPU and is incorporated into the SOP-GPU package. We profiled the computational performance of the SOP-GPU program with HI option for a range of biomolecular systems of varying size ($N = 10^1$–10^5 amino acids) to demonstrate that the program enables one to complete simulation runs in reasonable wall-clock time. We also compared the numerical accuracy of TEA-based treatment of HI against the results obtained with the Cholesky decomposition by performing Umbrella Sampling calculations of the unfolding free energy landscape for small proteins and by analyzing the unfolding kinetics of a generic multidomain protein.

Excellent agreement between the approximate and exact results has enabled us to carry out a comprehensive study of the influence of HI on a number of different processes, including the forced unfolding of a large multidomain protein, and the deformation of a biological filament. The results obtained illuminate the critical importance played by solvent-induced correlations influencing finite-rate dynamic processes involving biomolecules. Our results also necessitate the development of scalable computational modeling schemes for the accurate description of force generating properties of proteins, in which solvent-mediated many-body effects are accounted for. Utilizing these hydrodynamic interactions will allow better theoretical modeling of biological assemblies of all sizes to compare with both single-molecule force-based experiments as well as descriptions of biomolecular systems in the context of their cellular mechanical environment.

Methods

Hydrodynamic interactions

In Langevin Dynamics simulations in the overdamped limit, the x-, y-, and z-coordinates of N particles of the system in the three-dimensional space r_α, where index α runs through all the degrees of freedom for all N particles ($\alpha = 1, \ldots, 3N$), are propagated forward in time according to the Langevin equation of motion:

$$\Delta r_\alpha = \frac{\Delta t}{\gamma} (F_\alpha + \rho g_\alpha) \quad (1)$$

In eq. (1), $\rho = k_B T \sqrt{2/D_{\alpha\alpha}\Delta t}$ is the amplitude of random force, Δt is the timestep, $k_B T$ is the temperature, and γ is the friction coefficient, which is given by the Einstein-Stokes equation $\gamma = 6\pi\eta a$, with water viscosity $\eta = 7.0 \times 10^5$ pN ps/nm and the particle's size a. In eq. (1), $F_\alpha = -\partial U/\partial r_\alpha$ is the molecular force acting on the α-degree of freedom (U is the potential energy). In this approach, which ignores the hydrodynamic coupling of degrees of freedom, all particles are described by the same diffusion coefficient, $D = D_{\alpha\alpha} = k_B T/\gamma$, and stochastic kicks from solvent molecules are modeled using $N(0,1)$-distributed random forces g_α.

To account for solvent-mediated many-body effects, one can use an approach due to Ermak and McCammon.[25] In this approach, the equation of motion eq. (1) is transformed into the following equation:

$$\Delta r_\alpha = \sum_{\beta=1}^{3N} \frac{D_{\alpha\beta}}{k_B T} \Delta t F_\beta + \sqrt{2\Delta t} \sum_{\beta=1}^{3N} B_{\alpha\beta} g_\beta \quad (2)$$

In eq. (2), the hydrodynamic tensor $\mathbf{D}$, a real $3N \times 3N$ matrix in which an entry $D_{\alpha\beta}$ describes a contribution to the diffusion of degree of freedom α from the degree of freedom β. Tensor $\mathbf{D}$ can be represented by an $N \times N$ matrix of 3×3 submatrices $\mathbf{D}_{ij}$, each corresponding to a pair of particles i and j. In eq. (2), $\mathbf{B} = \mathbf{D}^{1/2}$; therefore, for the correct distribution of random forces a real $3N \times 3N$ matrix $\mathbf{B}$ must satisfy the condition $\mathbf{B}^\tau \mathbf{B} = \mathbf{D}$, where the superscript τ represents the transpose of a matrix. When $\mathbf{D}$ is diagonal with the equal matrix elements $D_{\alpha\alpha} = k_B T/\gamma$, we recover eq. (1).

The simplest explicit form of the hydrodynamic tensor $\mathbf{D}$ was first proposed by Oseen.[29] However, since this quantity is not always positively defined, the correct matrix $\mathbf{B}$ may not exist for some system's configurations. In this work, we use the Rotne–Prager–Yamakawa (RPY) tensor[30,31] which is a positive-definite quantity. The submatrices $\mathbf{D}_{ij}$ of RPY tensor are given by

$$\mathbf{D}_{ii} = \frac{k_B T}{6\pi\eta a} \mathbf{I} \quad (3a)$$

$$\mathbf{D}_{ij} = \frac{k_B T}{8\pi\eta} \left\{ \frac{1}{r_{ij}} \left[\left(1 + \frac{2a^2}{3r_{ij}^2}\right)\mathbf{I} + \left(1 - \frac{2a^2}{r_{ij}^2}\right)\frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2} \right] \right\} \quad \text{for } r_{ij} \geq 2a \quad (3b)$$

$$\mathbf{D}_{ij} = \frac{k_B T}{8\pi\eta} \left\{ \frac{1}{2a} \left[\left(\frac{8}{3} - \frac{3r_{ij}}{4a}\right)\mathbf{I} + \frac{r_{ij}}{4a}\frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2} \right] \right\} \quad \text{for } r_{ij} < 2a \quad (3c)$$

In eqs. (3a)–(3c), $\mathbf{I}$ is the identity matrix of rank 3 and a is the hydrodynamic radius of the particle. In this work, we use the same value of a for all particles.

Truncated expansion approximation

There are several methods for numerical computation of matrix **D**, including a fast multipole method[32] and Ewald summation method.[33] In this work, however, we utilized an approach in which the elements of the diffusion tensor **D** are computed on a pairwise basis. For each pair of residues i and j, the submatrix $\mathbf{D}_{ij}$ is computed directly according to eqs. (3a)–(3c). This approach has been extensively tested in conjunction with various algorithms for finding the matrix **B**,[27,28,34] and it does not require the periodic boundary conditions. Also, a straightforward way of computing **B** is Cholesky decomposition of **D**, but this method requires a total of $O(N^3)$ arithmetical operations and $O(N^2)$ memory space to store the whole matrix **D**. Clearly, such an approach is not most efficient for an implementation on a GPU, due to memory limitations. Several methods for approximate calculation of the matrix **B** have been developed, including the Fixman approximation,[26] Krylov subspace methods,[27] and the Truncated Expansion approximation.[28] In the Fixman method,[26] the square root of **D** is approximated using Chebyshev polynomials $p_L(\mathbf{D})$. Here, we only need a product $\mathbf{B} \cdot \mathbf{g} \approx p_L(\mathbf{D})\mathbf{g}$, which can be computed more efficiently than $p_L(\mathbf{D})$ itself. This method is iterative and requires estimation of the extreme eigenvalues of the matrix **D** for fast convergence. The Krylov subspace method[27] is similar to the Fixman method, but uses polynomials over Krylov subspace $\{\mathbf{g}, \mathbf{Dg}, \ldots, \mathbf{D}^{m-1}\mathbf{g}\}$. Although the use of Krylov subspace leverages the need to estimate the extreme eigenvalues, this method requires the whole matrix **D** as the input and several iterations to achieve convergence.

Therefore, in this work we employed the Truncated Expansion approximation (TEA).[28] This is by far the fastest method for calculation of the matrix **D**,[27] and it can be adapted to compute the matrix elements on a pairwise basis. The TEA method is less accurate than other methods available,[27,34] but given the coarse-grained nature of the SOP model, we accept this trade-off to achieve a significant computational speed-up, while also maintaining a reasonable level of accuracy. In the TEA approach, the matrix elements of **B** can be rewritten as $B_{\alpha\beta} = C_\alpha b_{\alpha\beta} D_{\alpha\beta}/\sqrt{D_{\alpha\alpha}}$, and eq. (2) can be recast as:[28]

$$\Delta r_\alpha = \frac{\Delta t}{\gamma} \sum_{\beta=1}^{3N} \frac{D_{\alpha\beta}}{D_{\alpha\alpha}} (F_\beta + C_\alpha b_{\alpha\beta} \rho g_\beta) \tag{4}$$

where

$$b_{\alpha\beta} = \begin{cases} 1, & \text{if } \alpha = \beta \\ b', & \text{if } \alpha \neq \beta \end{cases} \tag{5}$$

In eqs. (4) and (5), C_α and b' are given by

$$C_\alpha = \left(1 + \sum_{\alpha \neq \beta} b'^2 \frac{D_{\alpha\beta}^2}{D_{\alpha\alpha}D_{\beta\beta}}\right)^{1/2} \tag{6}$$

$$b' = \frac{1 - \sqrt{1 - [(N-1)\varepsilon^2 - (N-2)\varepsilon]}}{\sqrt{(N-1)\varepsilon^2 - (N-2)\varepsilon}} \tag{7}$$

where $\varepsilon = \langle D_{\alpha\beta}/D_{\alpha\alpha}\rangle$. Equations (5)–(7) allows one to parallelize the integration algorithm on a GPU. The memory demand

associated with using TEA scales with system size as $O(N)$, as opposed to $O(N^2)$ for Cholesky decomposition. The memory cost is small compared to the amount of memory required to store the lists of pairs of interacting particles.

Cholesky decomposition

We used the Cholesky decomposition method to assess the accuracy of the TEA method. Cholesky decomposition allows one to calculate the matrix elements of the diffusion matrix exactly. In this method, a decomposition of a positive-definite matrix **D** is employed to find the upper triangular matrix **B** so that the matrix **D** can be written as the product $\mathbf{D} = \mathbf{B}^\tau \mathbf{B}$. Numerically, this is achieved by solving a system of $N(N+1)/2$ linear equations. It takes $O(N^3)$ operations using the Cholesky–Banachiewicz or Cholesky–Crout algorithms to solve these equations, which differ only in their access pattern. We used the CPU implementation of Cholesky decomposition available in the GNU Scientific Library.[35] For the free energy calculations, we created our own GPU-based realization of Cholesky decomposition, which is now part of the SOP-GPU package.

SOP model

We employed the native topology-based SOP model of a polypeptide chain.[7,8] In the SOP model, each protein residue is represented by a bead, positioned at the residue's C_α-atom. The explicit form of the potential energy U_{SOP} expressed in terms of the coordinates of the C_α-atoms $\{r_i\} = r_1, r_2, \ldots, r_N$ of residues $1, 2, \ldots, N$ is given by

$$U_{SOP} = U_{FENE} + U_{NB}^{ATT} + U_{NB}^{REP} \tag{8}$$

In eq. (8), the first term is the finite extensible nonlinear elastic (FENE) potential, which describes the backbone chain connectivity:

$$U_{FENE} = -\sum_{i=1}^{N-1} \frac{k}{2} R_0^2 \log\left(1 - \frac{\left(r_{i,i+1} - r_{i,i+1}^0\right)^2}{R_0^2}\right) \tag{9}$$

where $k = 14$ N/m is the spring constant, and the tolerance in the change of the covalent bond distance is $R_0 = 2$ Å. The distance between the next-neighbor residues i and $i+1$, is $r_{i,i+1}$, and $r_{i,j+1}^0$ is its value in the native structure. The second term in eq. (8) takes into account the noncovalent (or nonbonded) interactions that stabilize the native state (structure). To describe these interactions, we use the Lennard–Jones potential:

$$U_{NB}^{ATT} = \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \varepsilon_h \left[\left(r_{ij}^0/r_{ij}\right)^{12} - 2\left(r_{ij}^0/r_{ij}\right)^6\right] \Delta_{ij} \tag{10}$$

In eq. (10), we assume that if noncovalently linked residues i and j ($|i-j| > 2$) are within the cut-off distance of 8 Å in the native state, then $\Delta_{ij} = 1$; and $\Delta_{ij} = 0$ otherwise. The value of ε_h quantifies the strength of the nonbonded (noncovalent)
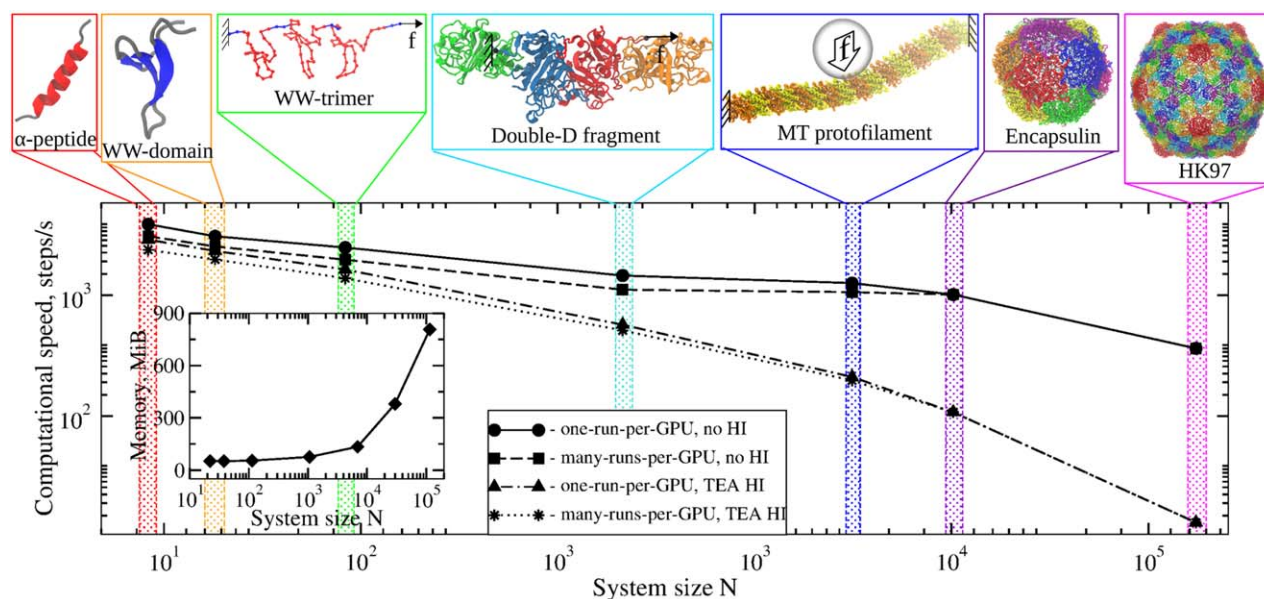
Figure 1. Computational performance of the SOP-GPU package with an approximate treatment of hydrodynamic interactions (TEA HI) and without hydrodynamics (no HI). The computational speed (number of integration steps per second) of an end-to-end realization of the SOP-GPU program implemented on a GPU Tesla K40 is estimated for a number of systems arranged in the order of their increasing size N (Table 1): i) α-peptide, ii) WW-domain, iii) trimer WW$_3$, iv) double-D fragment from fibrin, v) MT protofilament, vi) encapsulin shell, and vii) HK97 virus capsid. The computational speed is compared using: a) single-run-per-GPU approach without the HI (solid line) and with the TEA HI (dash-dotted line), and b) many-runs-per-GPU approach without the HI (dashed line) and with the TEA HI (dotted line). The *inset* plot shows the associated GPU memory usage vs. N for using the single-run-per-GPU approach (results of simulations with and without HI are indistinguishable on this scale).

interactions. In the SOP model, parameter ε_h sets the energy scale. The third term in eq. (8) accounts for the remaining non-native (nonbonded) interactions, which are treated as repulsive, including the bond angle between the residues i, $i + 1$, and $i + 2$:

$$U_{NB}^{REP} = \sum_{i=1}^{N-2} \varepsilon_l \left(\sigma_l / r_{i,i+1} \right)^6 + \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} \varepsilon_l \left(r_{ij}^0 / r_{ij} \right)^6 \left(1 - \Delta_{ij} \right) \quad (11)$$

where parameters $\varepsilon_l = 1$ kcal/mol and $\sigma_l = 3.8$ Å define the strength and range of the repulsion.

Model systems

The following model systems were used: (i) a fragment of an α-helical chain (α-peptide), (ii) a single WW-domain (WW), (iii) a trimer of linearly connected WW-domains (WW$_3$), (iv) a γ-γ-crosslinked double-D fragment from human fibrin(ogen) (double-D fragment), (v) a single microtubule (MT) protofilament, (vi) the nanocompartment encapsulin, and (vii) the bacteriophage HK97 capsid. The structures of these systems were taken from the Protein Data Base (PDB); see Figure 1. A summarized description of all the model systems, that is, system size, number of native contacts, and the total number of residue pairs for the calculation of diffusion tensor $\mathbf{D}$, is given in Table 1.

The α-peptide – a 22-residue peptide from the α-chain of human fibrinogen (residues 172–194[36];) was used in the calculation of the free energy landscape for the α-type protein. The WW-domain of mitotic rotamase Pin1,[37] a 34-residue long single protein domain, was used in the calculation of the free

energy for the β-sheet protein. *The WW-trimer WW$_3$*—used as a prototype of a multidomain protein, was constructed by connecting head-to-tail three identical WW-domains with glycine linkers, which were also attached to the N- and C-termini of WW$_3$. We used WW$_3$ to explore the effect of hydrodynamic interactions on protein forced unfolding. *The double-D fragment* from human fibrin(ogen)[38] consists of two identical D-regions from the abutted fibrin monomers connected through the γ-γ crosslinking sites. We used the double-D fragment to explore the influence of hydrodynamics on structural transitions in large-size protein assemblies. *The MT protofilaments* are long threads of longitudinally arranged α- and β-tubulin heterodimers (439 and 427 amino acids, respectively), which form long MT polymers in eukaryotic cells.[39] We used an 8-dimers long MT protofilament, employing nanoindentations *in silico*, to study the effects of hydrodynamic interactions on dissociation of longitudinal $\alpha\beta$-tubulin bonds. *The Encapsulin* and *Bacteriophage HK97* are examples of regular geometry biological protein assemblies.[40,41] We used encapsulin and HK97 (Head II state) to test the SOP-GPU (with HI option) capability (computational speed, memory usage).

Applications

Langevin dynamics

The biomolecular dynamics were obtained by performing numerical integration of the Langevin equations of motion for each particle's position $\mathbf{r}_i$ ($i = 1, 2, ..., N$) in the overdamped limit without HI [eq. (1)] and with HI [eq. (2)] using the first-order integration scheme. We implemented the Truncated

Table 1. Summarized description of and results of performance measurements for α-peptide, WW-domain, trimer WW₃, double-D fragment, microtubule (MT) protofilament, encapsulin shell, and bacteriophage HK97 capsid.

System	PDB code	System size N	Number of native contacts	Number of bead pairs	Optimal number of trajectories, s_{opt}
α-peptide	3GHG	22	32	992	350
WW-domain	1PIN	38	75	1406	170
trimer WW₃	1PIN	144	225	12,882	70
double-D fragment	1FZB	1062	3894	1,126,782	9
MT protofilament	1JFF	6928	21734	46,990,256	2
Encapsulin	3DKT	15,840	75138	890,813,562	1
bacteriophage HK97	1FT1	115,140	352224	13,257,104,460	1

Shown are the system size (number of amino acids), number of native contacts stabilizing the native state, number of pairs of beads (number of operations in the calculation of RPY tensor), and the optimal number of simulation runs on a GPU (Tesla K40) with the many-runs-per-GPU option.

Expansion based approximate treatment and the Cholesky decomposition based exact treatment of HI. The molecular forces were calculated from the total potential energy $U = U_{SOP}$ [eqs. (8)–(11)]. The Langevin equations were propagated with the time step $\Delta t = 0.08\tau_H \approx 20$ ps, where $\tau_H = \zeta \varepsilon_h \tau_L / k_B T$. Here, $\tau_L = (ma^2/\varepsilon_h)^{1/2} = 3$ ps, $\zeta = 50$ is the dimensionless friction constant for a residue in water ($\eta = \zeta m / \tau_L$), and $m \approx 3 \times 10^{-22}$ g is the average residue mass.[42] All the simulations with and without the HI were carried out at 300 K using the bulk water viscosity, which corresponds to the friction coefficient $\eta = 7.0 \times 10^5$ pN ps/nm. Numerical values of the hydrodynamic radius a vary in the literature from 1.5 Å[5] to 5.3 Å[1]. The TEA treatment of hydrodynamics correctly handles overlaps,[34] but the RPY tensor works better at describing the nonoverlapping particles. Since the $C_\alpha - C_\alpha$ distance in a polypeptide chain is 3.8 Å and due to self-avoidance, we set $a = 1.9$ Å.

Umbrella sampling simulations

To map the free energy landscape for unfolding of the WW-domain and α-peptide, we utilized the Umbrella Sampling technique (Supporting Information),[43] which is widely used to access the thermodynamics of biomolecules.[44–46] Using Umbrella Sampling requires generation and subsequent equilibration of initial conformations (windows), which span the entire range of reaction coordinate (elongation X). The potential of mean force was constructed using the Weighted Histogram Analysis Method.[44,47,48] To generate a set of initial structures for the α-peptide and the WW-domain, we ran 200 4-ms-long equilibration simulations for each molecule. We carried out force-ramp simulations by constraining the N-terminus and applying the pulling force to the C-terminus through a virtual cantilever spring $\kappa = 0.05$ kcal/mol/Å² ≈ 35 pN/nm moving with velocity $v_f = 2.5$ μm/s in the direction of X. We generated 1450 sampling windows of width 0.1 Å, corresponding to a 5 - 150 Å interval of X for the WW-domain; and 630 windows of width 0.1 Å, corresponding to a 25 - 88 Å interval of X for the α-peptide. These conformations were used in 10-ms equilibration runs. We carried out the Umbrella Sampling simulations without HI, and with HI using the exact (Cholesky) and approximate (Truncated Expansion) treatments.

Protein forced unfolding

In *force-clamp simulations*, which mimic force-clamp (e.g., AFM-based) measurements, the C_α-atom of the molecule at the N-terminus was constrained and the constant pulling force $\boldsymbol{f} = f\,\boldsymbol{n}$ with the magnitude f was applied to the C_α-atom at the molecule's C-terminus in the direction $\boldsymbol{n}$ coinciding with the end-to-end vector $\boldsymbol{X}$. In the force-clamp simulations for WW₃, the N-terminal C_α-atom in the first WW-domain was kept fixed whereas f was applied to the C-terminal C_α-atom of the third domain (Fig. 1). We generated 1000 runs for each force from $f = 100$ pN to 320 pN, with 20 pN increment. The unfolding time for each WW-domain was defined as the first time at which the end-to-end distance of the domain had exceeded 90 Å. *Force-ramp simulations* were carried out using the time-dependent force $f(t) = \kappa(v_f t - \Delta x)$, where v_f is the velocity of the cantilever base, represented by a virtual particle, κ is cantilever spring constant, and Δx is the displacement of the pulled residue from its initial position.[12] For double-D fragment, we constrained/pulled γCys139 in the first/second domain D (Fig. 1). We set the pulling velocity to either $v_f = 2.5$ or 25 μm/s, and the cantilever spring constant to $\kappa = 35$ pN/nm. The output was used to profile the unfolding force (F) as a function of molecular extension (X) or the *FX*-curves.

Nanoindentations *in silico*

For nanoindentations *in silico* on the MT protofilament, we harmonically connected the cantilever base with the spherical bead of radius R_{tip} mimicking the cantilever tip. The tip interacted with the particle via the repulsive Lennard–Jones potential:

$$U_{tip} = \sum_{i=1}^{N} \varepsilon_{tip} \left(\frac{\sigma_{tip}}{|r_i - r_{tip}| - R_{tip}} \right)^6 \qquad (12)$$

which produced an indentation on the particle's surface. In eq. (12), r_i and r_{tip} are coordinates of the i-th residue and the center of the tip, respectively; $\varepsilon_{tip} = 4.18$ kJ/mol, and $\sigma_{tip} = 1.0$ Å. For the cantilever tip-sphere (Fig. 6), we solved the following Langevin equation:

$$\gamma \frac{dr_{tip}}{dt} = -\frac{\partial U_{tip}(r_{tip})}{\partial r_{tip}} + \kappa((r_{tip}^0 - v_f t) - r_{tip}) \qquad (13)$$

where r_{tip}^0 is the initial position of the tip center, and the friction coefficient γ corresponds to the viscosity $\eta = 7.0 \times 10^6$

pN ps/nm. To generate the dynamics of a biological particle, we solved numerically eqs. (1) (without HI) or (2) (with HI), eqs. (8)–(11) for the particle, and eqs. (12) and (13) for the tip of radius $R_{tip} = 10$ nm. The cantilever base moving with constant velocity (v_f) exerted the compressive force $\boldsymbol{f} = f(t)\boldsymbol{n}$ in the direction $\boldsymbol{n}$ perpendicular to the particle surface; $f(t) = r_f t$ increased linearly in time t with the rate $r_f = \kappa v_f$ ($v_f = 3$, 10, and 30 μm/s and $\kappa = 50$ pN/nm). We monitored the tip position X, which defines the extent of deformation; the resisting force F was calculated using the energy output. To bend the MT protofilament, we constrained the C_α-atoms at the N-terminus of the first monomer (residues 248, 253, 257, 262, 325, 326, 329, 348, 349) and at the C-terminus of the last monomer (residues 98, 176, 177, 180, 221, 224, 225, 403, 407).

Results

GPU-based implementation of hydrodynamic interactions

The SOP-GPU with HI option can be run in the one-run-per-GPU mode (for large systems when N is comparable with the number of ALUs) and in the many-runs-per-GPU mode (for small systems when N is much smaller than the number of ALUs).[12] In our implementation of the TEA algorithm, all computations described in eqs. (3)–(7) are performed on a single GPU. First, we compute molecular forces F_β and random forces g_β for all particles. Next, we perform the numerical integration of the Langevin equations of motion using one thread of execution for each particle. At this stage, we compute the elements of tensor $\boldsymbol{D}$ for each particle by performing the summation over contributions from all particles [eq. (4)]. In this approach, each element of $\boldsymbol{D}$ is calculated twice, but we avoid synchronization issues and the need to store the whole matrix, which would be impossible given the relatively limited amount of memory in contemporary GPUs (up to 12 GiB). Here, the only operation performed on a CPU is the averaging of the off-diagonal elements of $\boldsymbol{D}$ [parameter ε in eq. (7)]. Also, because the positions of amino acid residues in 3D-space change gradually, parameter b' [eq. (7)] does not need to be updated frequently. In our implementation, we update b' every ~10–50 steps. More frequent updates do not improve accuracy and less frequent updates do not reduce significantly the computational time.

We also implemented the exact treatment of HI where we used the modified Cholesky–Banachiewicz algorithm, in which each thread computes a single row of the matrix $\boldsymbol{B}$. This requires that the whole matrix is processed by a single CUDA block (a group of threads sharing same memory) for synchronization purposes. In the developed implementation, we utilized shared memory for efficiency. While this approach works well when using the many-runs-per-GPU approach, which allows us to process simultaneously as many blocks as there are trajectories, it places an upper limit on system size due to the number of available threads in a single block: 341 residues on a GPU with CC 2.0 and later or 170 residues for CC lower than 2.0. Leveraging this would result in significant complica-

tion of the CUDA code, and simulations with $N > 10^2$ residues in centisecond timescales would not be feasible.

Computational performance

We assessed the computational performance of our implementation of the TEA algorithm on a GPU Tesla K40 (from NVIDIA) for the following systems: α-peptide, WW-domain, trimer WW$_3$, double-D fragment, MT protofilament, encapsulin shell, and HK97 capsid (Table 1). In the first set of benchmark tests, we performed long equilibrium simulations (10^7 integration steps) in the one-run-per-GPU mode for each system at $T = 300$ K with (TEA HI) and without (no HI) hydrodynamic interactions. We evaluated the computation speed (number of integration steps per second) and memory demand in the one-run-per-GPU mode. The profiles of computational speed and memory usage vs. system size N are compared in Fig. 1. When HI are switched off, the computational speed scales roughly linearly with N. This is because the computational speed is mostly determined by force calculations and by numerical integration of Langevin equations, both having the $O(N)$ complexity. Because the pair list generation, which has the $O(N^2)$ complexity, is performed infrequently (every 10^5 steps), it barely affects the runtime. Accounting for HI with the TEA approach results in the increase of algorithm complexity to $O(N^2)$ operations (Table 1). Since the integration of the Langevin equations is performed at each step, the runtime scales as ~N^2, which translates to a more rapid decrease of the computational time with N (Fig. 1). For small systems (10^1–10^3), the TEA HI simulations slow down several-fold. For large-size systems (~10^4–10^5), the runtime penalty is one-two orders of magnitude.

Next, we performed the same equilibrium simulations, but in the many-runs-per-GPU mode. There is the question of the optimal number of independent trajectories (s_{opt}) that can be run concurrently on a GPU. In principle, s_{opt} can be estimated using the formula:

$$s_{opt} \approx \frac{1}{N} \frac{M_{th} \cdot M_{cores}}{P}, \qquad (14)$$

which connects the total number of CUDA cores M_{cores}, the number of threads per GPU multiprocessor M_{th}, and the number of cores per GPU multiprocessor P. For Tesla K40 (based on Kepler architecture with compute capability 3.x), $M_{cores} = 2880$, $M_{th} = 512$, and $P = 192$. Hence, for the α-peptide ($N = 22$ residues) $s_{opt} \approx 350$. Because, eq. (14) provides only an estimate of s_{opt}, we performed equilibrium simulations described above in which we varied the number of trajectories s. Supporting Information Fig. S2 demonstrates that for small systems ($N = 10^1$–10^3 residues), the computational speed barely changes when s is small. This is because to fully load a GPU one needs to use up all the multiprocessors, which takes place when $M_{th} \cdot M_{cores}/P \approx 7500$ threads. For example, the computational speed is ~1.7×10^4–2.5×10^4 steps/second for the α-peptide when $s < 350$, ~1.3×10^4–1.7×10^4 steps/second for the WW-domain when $s < 170$, ~0.8×10^4–1.1×10^4 steps/second for WW$_3$ when $s < 70$, and ~0.3×10^4–0.5×10^4

steps/second for the double-D fragment when $s < 9$. We find that all the profiles have a similar initial slope, but after passing some optimal value $s = s_{opt}$ the computational speed starts decreasing more rapidly with s (Supporting Information Fig. S2). The GPU is now fully loaded and starts slowing down with each additional run. We used this crossover point to estimate s_{opt}. For large systems ($N = 10^4 – 10^5$), the GPU is loaded when $s \approx 1$, and the initial weak dependence on s is not observed.

The values of s_{opt} (Table 1) are the same for the simulations with and without HI, which shows that accounting for the hydrodynamic effects does not change the total numbers of threads of execution. The values of s_{opt} obtained were next used to carry out the equilibrium simulations (10^7 integration steps) in the many-runs-per-GPU mode to estimate the average computational speed. Figure 1 shows similar values of the computational speeds attained with the many-runs-per-GPU and one-run-per-GPU approaches, regardless of whether HI are accounted for or ignored. This means that the SOP-GPU program in the many-runs-per-GPU mode can be utilized to generate 10–100 independent runs with or without HI. The associated memory usage per trajectory for simulations with HI is very close to that without HI (*the inset* in Fig. 1), which means that biomolecular simulations with HI are not at all limited by the low memory of graphics cards.

Numerical accuracy

Unfolding energy landscape. To assess the numerical accuracy of TEA implementation, we profiled the unfolding free energy and unfolding force for the α-peptide and the WW-domain (Fig. 1) ΔG and F, as a function of the end-to-end distance X. Figure 2 compares ΔG vs. X and F vs. X obtained with no HI, with exact HI, and with TEA HI. Because HI should not influence the unfolding thermodynamics, the results of calculations of ΔG and F should not depend on whether HI are accounted for or ignored. We see that the profiles of ΔG vs. X (and F vs. X) for all three sets of simulations practically collapse onto the same curve both for the WW-domain and the α-peptide. Small deviations of the profiles of $\Delta G(X)$ and $F(X)$ obtained with TEA HI from the results obtained using exact HI are due to insufficient sampling (10 ms) of resulting conformations and the very slow but finite rate of change of X.

The WW domain represents a paradigm for describing folding and unfolding of the β-sheet proteins. The calculated profile of ΔG confirms that the WW domain is a two-state folder (Fig. 2a)[49] following the single-step kinetics of unfolding, F→U, from the native (folded) state F to the unfolded state U. By contrast, the profile of ΔG for the α-peptide shows that this is a downhill folder (Fig. 2b). We also calculated the average ΔG for unfolding of WW-domain and α-peptide from their folded state with $X_f = 2.7$ nm (WW-domain) and $X_f = 3.5$ nm (α-peptide) to the globally unfolded state with $X_u = 8.5$ nm (WW-domain) and $X_u = 6.8$ nm (α-peptide). For the WW-domain, $\Delta G = 40.9$ kcal/mol (no HI), 41.3 kcal/mol (exact HI), and 40.5 kcal/mol (TEA HI). For the α-peptide, $\Delta G = 24.0$ kcal/mol (no HI), 24.7 kcal/mol (exact HI), and 23.4 kcal/mol (TEA
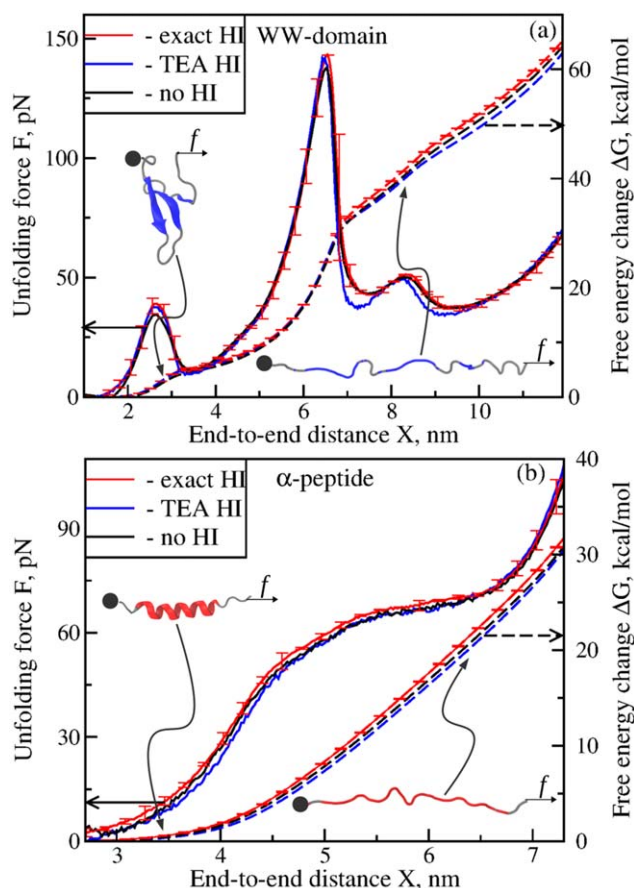
Figure 2. Thermodynamics of unfolding of the WW-domain (**panel a**) and the α-peptide (**panel b**); see Figure 1. Compared are the profiles of unfolding force F vs. end-to-end distance X (solid curves; left y-axes) and the unfolding free energy change ΔG vs. X (dashed curves; right y-axes) obtained with exact HI, TEA HI, and no HI treatments of hydrodynamics. Snapshots show the WW-domain and α-peptide in their folded and force-unfolded states.

HI). Hence, the results of exact and approximate treatments of hydrodynamics agree very well.

Unfolding kinetics. To explore the effects of hydrodynamic interactions on the kinetics of protein unfolding and to probe further the accuracy of TEA-based versus exact treatment of HI, we carried out force-clamp simulations for WW$_3$ (Fig. 1). Because in the trimer WW$_3$ any WW-domain can unfold at any given time with equal probability (domains are identical),[50] the unfolding kinetics can be characterized by analyzing the statistics of ordered unfolding time variates $\{t_{i:3}\}$ (order statistics), where $i = 1, 2, 3$.[51] We calculated the average first, second, and third unfolding times $t_{1:3}$, $t_{2:3}$, and $t_{3:3}$ using the results of simulations with exact HI, TEA HI, and no HI treatments of hydrodynamics. The profiles of order statistics data vs. f (Fig. 3) show that the average 1st, 2nd, and 3rd unfolding times all decrease with force. Surprisingly, the unfolding times from the simulations with exact HI and TEA HI are ~5–10 times shorter than the unfolding times from no HI simulations. Hence, hydrodynamic coupling accelerates unfolding of protein domains. Our results for ~$10^{-3}–10^{-5}$ s unfolding times (Fig. 3) also correlate well with the ~$10^3–10^5$ s^{-1} unfolding

rate constants from temperature jump and urea denaturation experiments.[52,53] The unfolding-time profiles obtained with exact HI and TEA HI practically overlap in the entire force range, implying a very good agreement between these two approaches (Fig. 3). Hence, the TEA treatment of HI offers an accurate description of the protein unfolding kinetics.

In our previous study, we showed that mechanical factors couple protein domains in multidomain proteins.[54] We estimated pair-wise correlations between the unfolding times of WW-domains in WW_3 t_I, $I = 1, 2, 3$ (parent data). Using statistically representative sets (1000 runs), we calculated for each IJ-pair of WW-domains ($I, J = 1, 2,$ and 3) the Pearson correlation coefficient:

$$p_{IJ} = \frac{\sum_n (t_I^n - \langle t_I \rangle)(t_J^n - \langle t_J \rangle)}{\sqrt{\sum_n (t_I^n - \langle t_I \rangle)^2}\sqrt{\sum_n (t_J^n - \langle t_J \rangle)^2}} \tag{15}$$

and the Spearman rank correlation coefficient:

$$s_{IJ} = \frac{\sum_n (r_I^n - \langle r_I \rangle)(r_J^n - \langle r_J \rangle)}{\sqrt{\sum_n (r_I^n - \langle r_I \rangle)^2}\sqrt{\sum_n (r_J^n - \langle r_J \rangle)^2}} \tag{16}$$

In eqs. (15) and (16), t_I^n is the unfolding time of the Ith domain observed in the nth trajectory, r_I^n is the rank of t_I^n, respectively, and $\langle \ldots \rangle$ denotes ensemble averaging. The force profiles of p_{IJ} and s_{IJ} ($-1 < p_{IJ}, s_{IJ} < 1$) are displayed in Figure 4. We observe very strong negative correlations between the unfolding transitions in the 1st WW-domain (constrained) and 2nd domain (in the middle) reflected in large negative values of $p_{12} = -0.3$–0.7 and $s_{12} = -0.3$–0.6, and weak negative correlations between the unfolding transitions in the 1st and 3rd WW-domain (pulled) corresponding to small negative values of $p_{13} = s_{13} = -0.05$–0.15, and $p_{23} = -0.05$–0.2 and $s_{23} = -0.05$–0.15 in the 150–300 pN force range (Fig. 4). Negative correlations between domains I and J imply that the longer (shorter) unfolding times for domain I correspond to the shorter (longer) unfolding times for domain J on average. These very similar results correspond to exact-HI and TEA-HI simulations. In contrast, no-HI simulations predict that correlations between the unfolding transitions in the 1st and 2nd domains change from large negative ($p_{12} = s_{12} = -0.3$–0.6 for $f = 100$–175 pN) to large positive ($p_{12} = 0.1$–0.4 and $s_{12} = 0.3$–0.6 for $f = 175$–300 pN), and that correlations between the unfolding times for the 1st and 3rd domains and 2nd and 3rd domains are negligible in the 100–250 pN force range but weak and positive for $f > 250$ pN (Fig. 4). Positive correlations imply that the longer (shorter) unfolding times for domain I correspond to the longer (shorter) unfolding times for domain J on average. Hence, the results obtained demonstrate that biomolecular simulations with HI ignored incorrectly describe dynamic coupling of protein domains in multidomain proteins.

Unfolding pathways. We also analyzed the unfolding pathways. A pathway I-J-K is a sequence of unfolding transitions, in which the WW-domains I, J, and K have unfolded first, second, and third, respectively. The pathway probabilities are accumu-
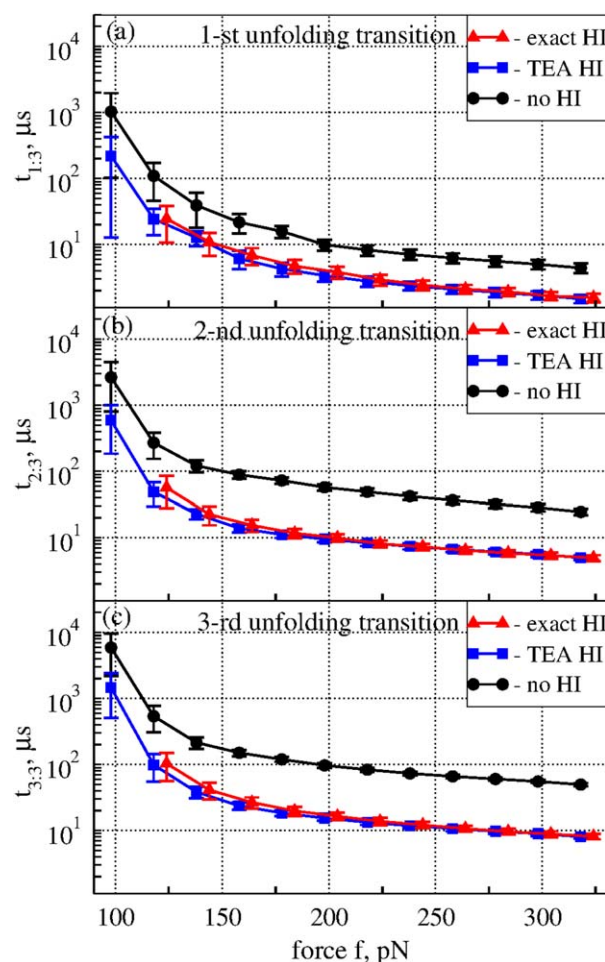
Figure 3. Order statistics of the forced unfolding times from force-clamp simulations for WW_3 (Fig. 1; Table 1). The average unfolding times for the 1st, 2nd, and 3rd unfolding transition $t_{1:3}$ (**panel a**), $t_{2:3}$ (**panel b**), and $t_{3:3}$ (**panel c**) vs. constant force f. Compared are the results of simulations with no HI (black circles), exact HI (red triangles), and TEA HI (blue squares). Averages and standard deviations are calculated over 10^3 trajectories (data for exact-HI simulations are shifted along the x-axis for clarity). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary. com.]

lated in Table 2. The exact-HI and TEA-HI simulations display very similar pathway probabilities, which differ from probabilities for no-HI simulations. For example, the no-HI simulations predict that 3-2-1 and 3-1-2 are the most dominant pathways in the 100-280 pN force range. For $f = 100$ pN, the pathways 3-2-1 and 3-1-2 are detected, respectively, 30.4% and 23% of the time. Yet, the exact-HI and TEA-HI simulations display comparable pathway probabilities for all pathways, that is, ~14–23% (exact-HI) and ~11–18% (TEA-HI). Although at higher forces pathways 3-2-1 and 3-1-2 become increasingly more dominant in all three simulation schemes, at $f = 220$ pN, simulations with HI predict 80% (pathway 3-2-1) vs. 20% (pathway 3-1-2) for the exact-HI and 83.6% (pathway 3-2-1) vs. 16.4% (pathway 3-1-2) for the TEA-HI. At this 220 pN force, the no-HI simulations predict 99% (pathway 3-2-1) vs. 1% (pathway 3-1-2). Hence, in terms of pathway probabilities, the results of simulations with and without HI disagree in the entire 100–280 pN force range. This can also be seen in the scatter plots in
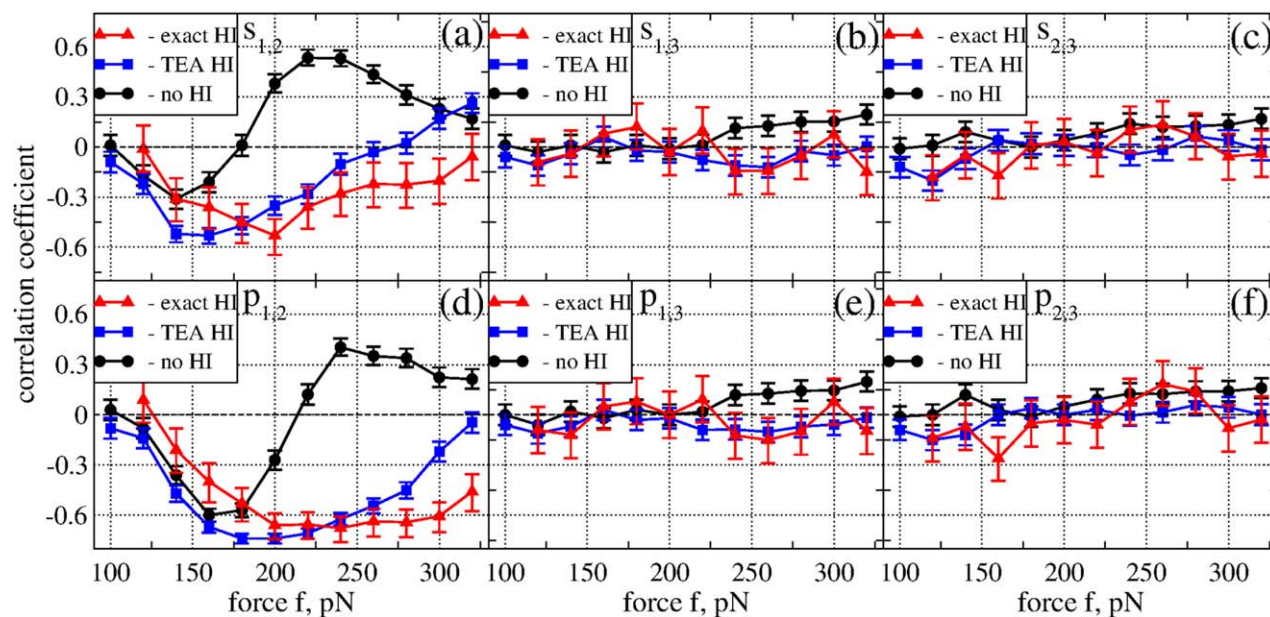
Figure 4. Pairwise correlations characterizing dynamic coupling of the unfolding transitions in WW-domains (Fig. 3) from pulling simulations for WW_3 with constant force f (Fig. 1; Table 1). Shown are the profiles of Spearman rank correlation coefficient s_{IJ} (**panels a–c**) and Pearson correlation coefficient p_{IJ} (**panels d–f**), $I \neq J = 1, 2, 3$, calculated using the unfolding times for the 1st and 2nd WW domains (s_{12} and p_{12}), 1st and 3rd domains (s_{13} and p_{13}), and 2nd and 3rd domains (s_{23} and p_{23}). The results are compared for simulations with no HI (black circles), exact HI (red triangles) and TEA HI (blue squares) treatment of hydrodynamics (error bars indicate 95% confidence interval). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Supporting Information Figs. S3 and S4, which demonstrate that the TEA-HI simulations better represent the unfolding pathways not predicted in the no-HI simulations. Hence, the neglect of hydrodynamic interactions leads to inaccurate descriptions of unfolding pathways.

Forced unfolding of double-D fragment

The very good agreement we obtained comparing the results of exact vs. approximate treatments of hydrodynamics for WW_3, has enabled us to explore the effect of hydrodynamic interactions on the dynamics of forced unfolding of a large protein complex–the double-D fragment (Fig. 1). We demonstrated in our previous study[15] that the noncovalent D-D interface formed by two neighboring fibrin monomers exhibit complex multistep unfolding transitions. We performed the force-ramp simulations of double-D fragment with HI (TEA HI) and without HI (no HI) and compared these results obtained with slow and fast cantilever velocities $v_f = 2.5$ and 25 μm/s. The force (F)-extension (X) profiles (Fig. 5) show that the corresponding FX curves differ only in minute details under slow force loading, but significant differences become manifest under faster loading. Interestingly, the FX-spectra obtained with HI are more sawtooth-like, that is, have more pronounced force peaks with higher peak-forces and deeper valleys, and show much lower tension noise level and a lower lying baseline. The FX curves from TEA-HI simulations with 2.5 and 25 μm/s cantilever velocity have the baseline running 10–20 pN and 100–200 pN lower, respectively, than the FX curves from no-HI simulations. We also performed structure analysis using the output from TEA-HI and no-HI simulation, and found no

substantial differences in the unfolding mechanisms. This can also be learned from the similar QX-profiles (Fig. 5).

The disruption of the D-D interface induces a conformational transition from the "closed" state with hidden D-D junction to the "open" state with the D-D interface fully disrupted (see the first small $\sim$35–40 pN force peak at the $\sim$12 nm separation distance in Fig. 5a). The following six high-amplitude force peaks correspond to sequential unraveling of the γ nodules in the globular D regions (three force peaks for each D region).[15] Under fast loading (25 μm/s), the last two transitions are resolved in the TEA-HI simulations and are not

Table 2. Unfolding pathways for the trimer WW_3.							
		Pathway probability, %					
Force, pN	HI	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1
100	exact HI	19.0	14.3	14.3	23.8	11.9	16.7
160	exact HI	0.0	1.0	0.0	5.6	32.4	61.0
220	exact HI	0.0	0.0	0.0	0.1	18.7	81.2
280	exact HI	0.0	0.0	0.0	0.0	5.9	94.1
100	TEA HI	11.0	18.3	15.1	16.6	21.6	17.4
160	TEA HI	0.0	0.9	0.0	2.3	36.0	60.8
220	TEA HI	0.0	0.0	0.0	0.0	16.4	83.6
280	TEA HI	0.0	0.0	0.0	0.0	4.4	95.6
100	no HI	7.8	11.0	8.5	19.3	23.0	30.4
160	no HI	0.0	0.0	0.0	0.0	20.2	79.8
220	no HI	0.0	0.0	0.0	0.0	1.0	99.0
280	no HI	0.0	0.0	0.0	0.0	0.0	100.0

Shown are the probabilities of observing a particular sequence of unfolding transitions in WW-domains calculated from the force-clamp pulling simulations with exact HI, TEA HI, and no HI treatments of hydrodynamic interactions for each value of the constant force $f = 100$, 160, 220, and 280 pN.
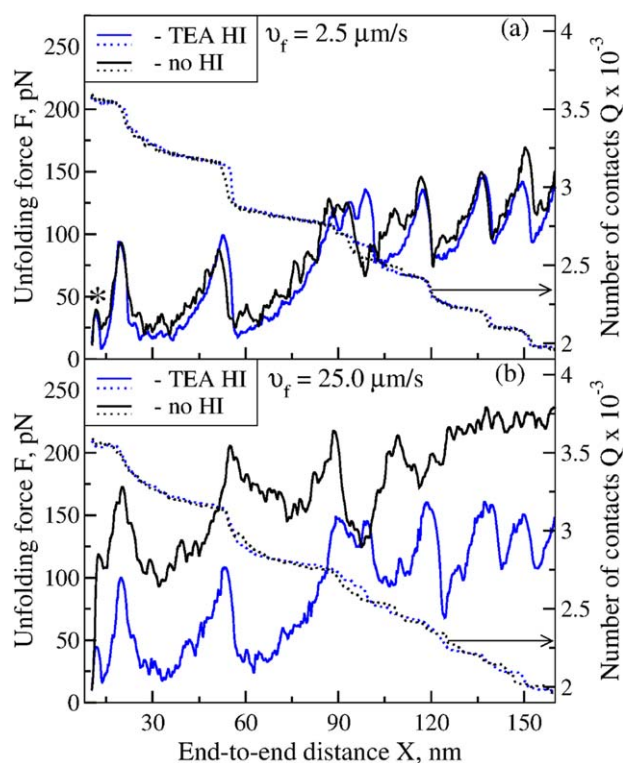
Figure 5. The unfolding transitions in double-D fragment (Fig. 1; Table 1). Shown are the unfolding force spectra, that is, force (*F*)–extension (*X*) curves (left *y*-axes), and the profiles of native contacts *Q* vs. *X* (right *y*-axes) from the force-ramp simulations with the pulling force increasing with $v_f = 2.5$ μm/s (**panel a**) and 25 μm/s (**panel b**). We used the cantilever spring constant $\kappa = 35$ pN/nm. Compared are the *FX*-and *QX*-curves from simulations with no HI (black curves) and TEA HI (blue curves). In **panel a**, the first small ~40 pN force peak at the ~12 nm distance corresponds to the dissociation of D-D interface (marked by the asterisk). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

resolved in the no-HI simulations (Fig. 5b). Hence, accounting for the HI makes the *FX* profiles look sharper and more experimental-like, in which tension fluctuations are much reduced. Yet, the inclusion of HI does not alter the unfolding pattern.

Bending deformation of MT protofilament

Next, we explored the influence of HI on the dynamics of bending deformations of biological particles using a short 8 $\alpha\beta$-tubulin dimer long microtubule (MT) protofilament (Fig. 1). We performed *in silico* nanoindentations of the MT protofilament with and without HI and compared the results obtained with slow, intermediate, and fast cantilever velocities $v_f = 3, 10,$ and 30 μm/s (see Supporting Movie S1).

The force (*F*)-deformation (*X*) spectra (*FX* curves) are displayed in Figure 6. The differences in the lineshapes obtained with and without HI grow with the loading rate and become significant under fast loading. At $v_f = 3$ μm/s, the inclusion of HI only slightly affects the values of critical force *F** and critical distance *X** (Fig. 6), that is, $X^* \approx 6.3$ nm and $F^* \approx 0.19$ nN (no-HI) vs. $X^* \approx 7.3$ nm and $F^* \approx 0.21$ nN (TEA-HI). Here, the critical force corresponds to the largest deformation before the breakage occurs (snapshots 2 and 3 in Fig. 6a). With the inclu-

sion of HI, the force peak becomes sharper and the baseline is less noisy and runs lower. Increasing v_f to 10 μm/s barely changes the *FX* spectrum for the TEA-HI simulations in terms of the lineshape and the values of *F** and *X**, but results in a huge line-broadening in the spectrum for the no-HI simulations (Fig. 6b). Under fast loading at $v_f = 30$ μm/s, the force peak barely shifts in the spectrum from the TEA-HI simulations, but the spectrum from the no-HI simulations becomes featureless (Fig. 6c). We extracted the average initial slope *dF/dX*, which was used to calculate the flexural rigidity *EI* (see Supporting Information). We found that $dF/dX = 30.6$ pN nm (TEA HI) vs. 31.1 pN nm (no HI), and that $EI = 20,500$ pN nm^2 (TEA HI) vs. 43,800 pN nm^2 (no HI). Hence, the neglect of hydrodynamic interactions leads to a two-fold overestimation of the flexural rigidity of the MT protofilament.
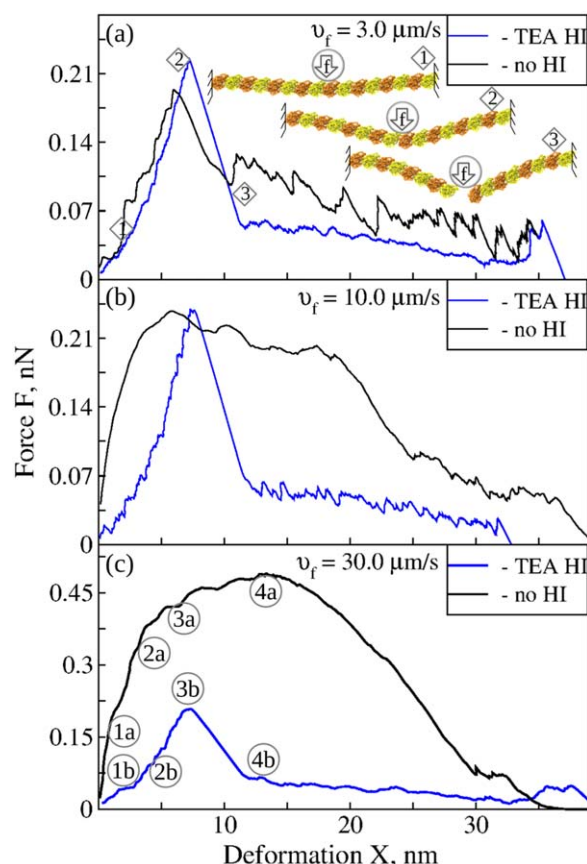
Figure 6. The dynamics of deformation of MT protofilament (Fig. 1; Table 1). Shown are the force (*F*)–deformation (*X*) curves obtained from *in silico* nanoindentations with compressive force increasing with $v_f = 3$ μm/s (**panel a**), 10 μm/s (**panel b**), and 30 μm/s (**panel c**). The cantilever spring constant is $\kappa = 50$ pN/nm, and the cantilever tip radius is $R_{tip} = 10$ nm. Compared are the *FX*-spectra from simulations with no HI (black curves) and TEA HI (blue curves). In **panel a**, snapshots corresponding to the similarly numbered regions 1, 2, and 3 in the *FX*-spectra show the deformation progress, from the bent structure 1 (ascending part of the *FX*-spectrum), to the bent structure 2 before the MT protofilament dissociation (force maximum in the *FX* spectrum), and to the disrupted structure 3 (local minimum in the *FX* spectrum). In **panel c**, the locations on the *FX*-curves numbered 1–4 correspond to the snapshots displayed in Figure 7. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To probe the origin of these differences, we performed the principal component analysis (see Supporting Information) using the structure output from the TEA-HI and no-HI nanoindentations of the MT protofilament under fast force loading with $v_f = 30$ μm/s (Fig. 6c). The first mode, which accounts for ~83% of bending deformation, corresponds to the out-of-plane displacement of the MT protofilament beam in the direction of applied force (see Supporting Information Movie S2 and Fig. 6a), but the extent of bending is larger in the presence of HI. Hence, hydrodynamic interactions couple the displacements of relevant elements of the particle's structure. The increased amplitude of these collective displacements corresponds to the decrease of the deformation force (Fig. 6c).

Discussion

Simplified models of biomolecules allow researchers to perform meaningful simulations on biologically relevant time scales.[55] The SOP model of biomolecules (RNA and proteins)[7,8] is a native topology based coarse-grained model, which, in conjunction with the computational acceleration available on a GPU,[12,13] enables one to follow the dynamics of biomolecular systems in the relevant experimental subsecond timescale (Supporting Information Fig. S1). To harness the computational power of graphic processors, the SOP-GPU package utilizes the CUDA Toolkit (from NVIDIA). The SOP model has already been used by many research groups in the computational exploration of a broad range of biological systems.[13,15–18,50,51,54,56–60]

There are limitations of the SOP model as with any model: (i) the atomic details are ignored; (ii) only native contacts are accounted for; (iii) there are no side chains (in the C_α-model); and (iv) electrostatic interactions are not described explicitly. Conversely, laser optical tweezers and atomic force microscopy cannot resolve structural transitions on the sub-nanometer lengthscale; the 10–20 nm indenter size used in nanomanipulations with AFM makes the atomic structure details irrelevant. Although unfolding of proteins occurs via sequential unraveling of the tertiary and then secondary structure blocks formulating the native topology, in mechanical deformation of whole biological particles, the overall geometry of the particle's structures and specific arrangements of large building blocks (e.g., capsomers in virus shells and protofilaments in MT polymers) play an important role. Hence, it is sufficient to model only contacts that reinforce the native structure of protein domains. Although the side chains are not included into the SOP model, their influence is implicit in the long 8 Å cut-off distance [eqs. (10) and (11)]. Also, the explicit treatment of side chains is feasible with the C_α-C_β version of the SOP model, and the inclusion of electrostatic interactions is possible with the screened Coulomb potential.[61] Notwithstanding the limitations of the SOP model summarized above, the model has the following major advantages. The simplicity of the SOP model allows using the relevant experimental force-loading rate conditions to describe the biomolecular transitions in the diffusion-controlled limit. The simple form of the energy function [eqs. (8)–(11), which involves the potential energy terms

also used in more detailed descriptions [e.g., FENE potential; see eq. (9)], makes it possible to implement the SOP model on a GPU. Owing to a huge computational acceleration attainable on a GPU, this opens a new direction of research of the force-generating properties of entire biological particles (viruses, cellular organelles, filaments, etc.)[15–18,55]

To account for solvent-mediated many-body effects in Langevin dynamics schemes, here, we have employed the Truncated Expansion approximation for calculation of the Rotne–Prager–Yamakawa hydrodynamic tensor **D**. We assessed the computational performance of the end-to-end realization of Langevin dynamics with HI on a GPU by profiling the computational speed and memory demand (Fig. 1 and Supporting Information Fig. S2) for a range of model systems ($N = 10^1$–10^5 residues; Table 1). Benchmark testing showed that these systems can be computationally described with the HI in reasonable wall-clock time and that these simulations are not limited by the low memory of graphics cards. To generate 1 ms of Langevin dynamics (with time step of 20 ps) with TEA HI on a single GPU Tesla K40 using the many-runs-per-GPU mode, it would take ~2 h for the D-dimer ($s_{opt} = 9$); and ~43 hours for the 8-dimers long MT protofilament ($s_{opt} = 2$). It takes 5–20 times longer, depending on system size, to simulate the dynamics with HI (Fig. 1), but the global transitions in biomolecules occur on a faster timescale in the presence of HI (Figs. 3 and 5, and 6, Supporting Information Figs. S3 and S4). For example, in the force-clamp simulations for the WW-domain (Fig. 3), the three-fold computational slowdown is more than balanced by the approximately four-fold speed-up of unfolding kinetics.

We assessed the numerical accuracy of our GPU-based realization of the TEA approach to describing the HI by comparing the results of free energy calculations and protein forced unfolding simulations generated with the approximate and exact treatments of hydrodynamics. The profiles of unfolding free energy for the TEA-HI and exact-HI treatments for the WW-domain and the α-peptide practically collapse onto the same curve (Fig. 2). The thermodynamic state functions (ΔG) characterizing forced unfolding in these model proteins estimated based on these two approaches agree very well. The small size of the trimer WW$_3$, used as an example of a multidomain protein, allowed us to directly compare how simulations with the exact and approximate treatments of HI describe the protein unfolding. Analysis of the unfolding kinetics for WW$_3$ revealed overlapping profiles of ordered unfolding times (Fig. 3), as well as Pearson correlation and Spearman rank correlation coefficients (Fig. 4), and the very similar values of pathway probabilities (Table 2) obtained with the exact and TEA treatments of HI. These results imply an almost quantitative agreement between the results of exact vs. TEA treatments of HI. Hence, the TEA approach offers a convenient, yet accurate description of the thermodynamics and kinetics of protein unfolding.

The main results from this part of our studies are the following. Hydrodynamic interactions: (i) accelerate the kinetics of unfolding transitions in protein domains in multidomain proteins and polyproteins (Fig. 3 and Supporting Information Figs.
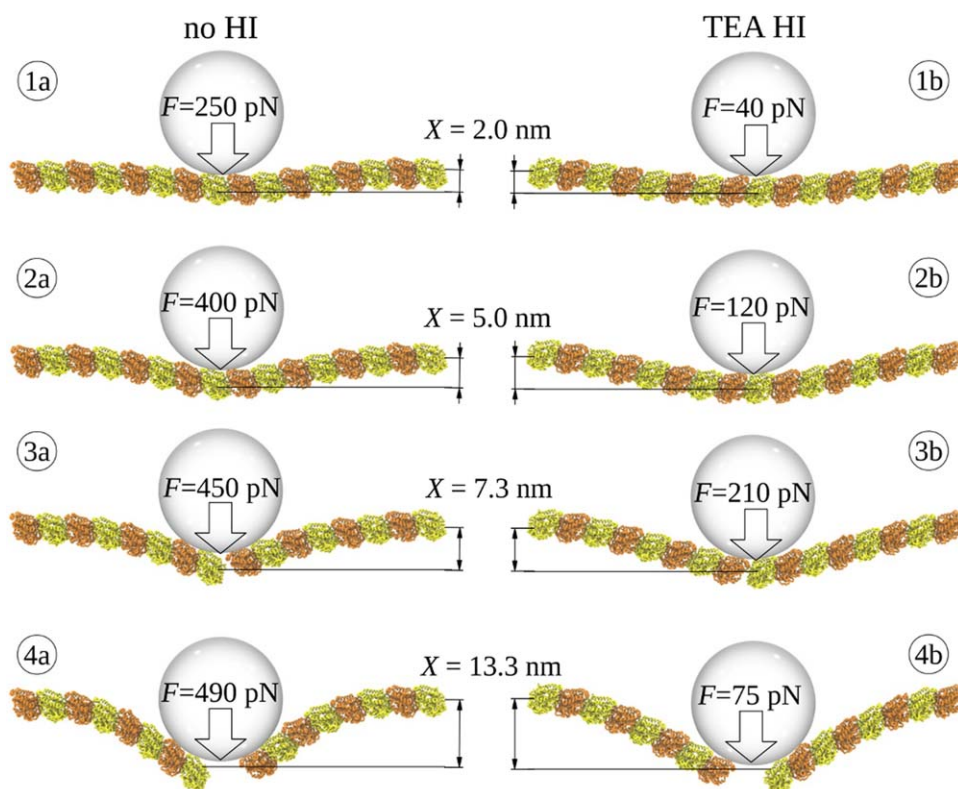
Figure 7. Structural snapshots explaining detectable differences in the dynamics of deformation of the 8 $\alpha\beta$-dimers long MT protofilament fragment (Table 1) in the absence of HI (no HI; structures 1a, 2a, 3a, and 4a on the left) and in the presence of HI (TEA HI; structures 1b, 2b, 3b, and 4b on the right). The snapshots for $X = 2$, 5, 7.3, and 13.3 nm deformation correspond to the *FX*-curves in Figure 6c. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

S3 and S4) in agreement with experiments; (ii) couple these transitions when the tension is sufficiently high ($>$ 160–200 pN for WW-domain; Fig. 4), thereby making them interdependent in agreement with our previous theoretical study[54]; and (iii) influence the kinetic partitioning of unfolding transitions into various pathways (Table 2).

These results can be understood by considering the dynamics of tension propagation in a polypeptide chain together with the kinetics of protein forced unfolding. When the pulling force is small, that is, it barely exceeds the threshold force for unfolding, the timescale of tension propagation and redistribution is much shorter than the unfolding kinetics. Under these conditions, the effect of hydrodynamic interactions between residues in the same or different WW-domains is negligible, and they unravel nearly independently, with (almost) equal probability regardless of their position in the trimer WW$_3$. Hence, any of the six pathways (1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2, 3-2-1) should be observed with the equal probability of 1/6. When the pulling force is large, the hydrodynamic effects become important. Because HI are a kinetic effect, that is, they couple the displacements of various residues, the native contacts stabilizing the WW-folded state disrupt roughly at the same time. The larger the force, the more time-coordinated are the dissociations of native contacts. In a sense, the WW-domains become destabilized kinetically, which shortens their lifetime and accelerates the unfolding rate (Fig. 3), but not thermodynamically. Recall that the WW-domains are equally

stable thermodynamically with and without HI (Fig. 2). At large force, the timescale of tension propagation becomes comparable with the unfolding kinetics, and the domains closer to the point of force application (i.e., the 3rd and 2nd WW-domains) unfold more rapidly. This is the cause of kinetic partitioning and emergence of dominant pathways 3-2-1 and 3-1-2 at large forces (Table 2). However, because tension propagation is more rapid and tension redistribution is more uniform in the presence of HI, the kinetic partitioning becomes stronger at larger forces (compared to when HI are off) when the unfolding kinetics exceeds the rate of tension propagation.

The excellent agreement we obtained between the results of approximate and exact treatments of HI, has enabled us to explore the influence of hydrodynamic coupling on dynamic processes involving large biomolecular complexes comprised of a number of protein building blocks (Table 1). We studied the dynamics of protein forced unfolding for the D-dimer and the dynamics of deformation of an MT protofilament (Fig. 1; Table 1). The results from this part of our studies showed that the inclusion of HI: (iv) modulates the deformation dynamics of a biological particle, which become more deterministic (i.e., less variable) with the reduced role of tension or stress fluctuations; (v) results in the force-extension/deformation spectra (*FX*-curves) that are drastically different from the corresponding *FX*-curves collected in the absence of HI, especially in the limit of fast loading (Figs. 5 and 6). However, (vi) the HI do not change the physical picture underlying the unfolding/deformation. The *FX*-curves
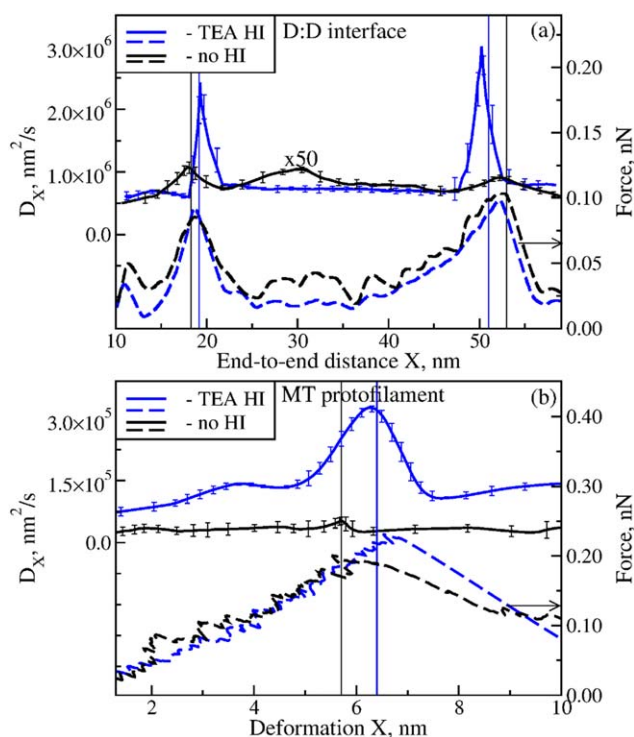
Figure 8. Profiles of the position-dependent diffusion coefficient D_X (left y-axis) vs. X for the double-D fragment unfolding (**panel a**), and the MT protofilament deformation (**panel b**) based on experiments *in silico* with TEA HI (blue curves) and with no HI (black curves). Also shown are the corresponding FX-curves from Figures 5a, and 6a (dashed curves; right y-axis), respectively, for the D-dimer (**panel a**), and MT protofilament (**panel b**). For the D-dimer, we only display the initial 60 nm portions of the D_XX- and FX-curves. Thin vertical lines correlate the critical values of reaction coordinate X in FX- and D_XX-curves marking the global transitions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

obtained using the TEA treatment of hydrodynamics (Figs. 5) look much more similar to the typical experimental FX-spectra[15] compared to the results obtained with the no-HI approach. The force peaks, which quantify the limits of the system's stability to the applied pulling/compressive force, are more distinguishable and they are now reminiscent of the experimental sawtooth-like force signals due to the lower running baseline and weaker noise level. The lower lying baseline in the FX-spectra means that faster tension (or stress) propagation and a more uniform tension (or stress) distribution exists in the D-dimer (MT protofilament); see Figure 5 (Fig. 6). This leads to a smaller unfolding (deformation) force reading for the D-dimer (MT protofilament). Weaker noise levels imply reduced mechanical perturbations experienced by the system due to the external tension/stress factor. This is because hydrodynamic interactions couple structural elements, thereby making them able to undergo correlated displacements and essentially mask the mechanical fluctuations. These correlated displacements also echo with the coupled unfolding transitions observed in the trimer WW$_3$ in the presence of HI that we discussed above. Stated differently, in the presence of HI the displacements are less individual (or local) and more collective (or global). Hence, the inclusion of HI accelerates tension or stress propagation and/or redistribution, and tension or stress fluctuations are much reduced.

Surprisingly, the inclusion or neglect of HI does not change qualitatively the mechanism of global transitions (unfolding, deformation, and breakage). This conclusion is uniformly consistent with the results of structural analyses for the D-dimer unraveling, and MT protofilament deformation and breakage. The most striking evidence supporting this notion comes from the principal component analysis of nanoindentations of the MT protofilament. When the HI are switched on, the out-of-plane deformations of the $\alpha\beta$-tubulin dimers at the clamped ends and around the indenter (first principal mode) become strongly coupled (Supporting Information Movie S2), which makes these displacements time-coordinated. This explains the lower lying FX-curves (Fig. 6), and smaller resulting values of the deformation force for a more bendable beam structure. This last point can be better understood by comparing the structures of MT protofilament and deformation forces (F) for the same values of deformation $X = 2$, 5, 7.3, and 13.3 nm compared in Figure 7 for fast force loading ($v_f = 30$ μm/s) with and without the HI. We observe the very different values of deformation forces: $F = 250$ pN (no HI) vs. 40 pN (TEA HI) for $X = 2$ nm (weakly bent conformation); $F = 400$ pN (no HI) vs. 120 pN (TEA HI) for $X = 5$ nm (weakly bent conformation); $F = 450$ pN (no HI) vs. 210 pN (TEA HI) for $X \approx 7$ nm (strongly bent conformation); and $F = 490$ pN (no HI) vs. 75 pN (TEA HI) for $X \approx 13$ nm (dissociated state). Taking into account the $\sim$64 nm size of 8 $\alpha\beta$-tubulin dimers long MT protofilament (Fig 1), we can appreciate the fact that in the presence of HI the displacements of the relevant elements of MT protofilament structure are coordinated on a truly global scale.

Our finding that the neglect or inclusion of HI does not alter the mechanism of transitions correlates well with our conclusion that hydrodynamic coupling is entirely a kinetic effect, which mainly correlates the residues undergoing concerted displacements. To illustrate this point, we used an approach described in Ref. [62] to calculate the local (position-dependent) diffusion coefficient D_X along the reaction coordinate X,

$$D_X = D(\langle X \rangle) = \frac{\text{var}(X)}{\tau_X} \tag{17}$$

In eq. (17), $\langle X \rangle$ is the average value of X, that is, the end-to-end distance for the D-dimer (Fig. 5) and deformation for the MT protofilament (Fig. 6), $\text{var}(X) = \langle X^2 \rangle - \langle X \rangle^2$ is the variance, and $\tau_X = \int_0^\infty \langle \delta X(t) \delta X(0) \rangle dt / \text{var}(X)$ is the characteristic time ($\delta X(t) = X(t) - \langle X \rangle$). The results of calculation of D_X for the D-dimer and MT protofilament are presented in Figure 8. Both for simulations with and without the HI, the profiles of D_X and F vs. X are peaked around the critical values of X, which mark these global transitions: unfolding events in the D-dimer; and breakage of the MT protofilament. The higher lying baseline and the peaks of D_X correspond to simulations with HI (Fig. 8). Hence, the large-amplitude global transitions are accelerated many-fold when the hydrodynamic interactions are switched on.

Conclusion

Notwithstanding the huge computational speed-up attained when Langevin simulations are ported to a GPU,[12,13] it is not yet possible to describe the hydrodynamic interactions exactly

in reasonable wall-clock time, even for a relatively small system of $N = 10^3$ residues (D-dimer; Table 1) and even at the one-bead-per-residue level. This situation is due primarily to pro-hibitively expensive numerical costs associated with the com-putation of the Rotne–Prager–Yamakawa hydrodynamic tensor. Here, we implemented an algorithm for fast yet accurate cal-culation of the hydrodynamic tensor on a GPU, which is based on the Truncated Expansion approximation.[28] This develop-ment allowed us to carry out a comprehensive study of the influence of hydrodynamic interactions on the biomechanical and force-generating properties for a range of biomolecular systems from small proteins to large-size protein assemblies.

Biomolecular simulations of protein forced unfolding with the hydrodynamic interactions neglected lead to an inaccurate description of the unfolding kinetics and pathways and a signifi-cant underestimate of the role of dynamic coupling of protein domains forming a multidomain protein. The visual differences in the force-extension/deformation spectra and corresponding size-able disagreement, for example, in the values of slope and criti-cal forces for unfolding/deformation can potentially lead to erroneous estimation of the physico-chemical and thermody-namic characteristics. These problems can be alleviated using TEA treatment of hydrodynamics. Undoubtedly, hydrodynamic interactions represent a more accurate physical picture of the properties of protein domains imparted by solvent to biomole-cules undergoing structural transformation, whether this is as part of their functional mechanism in the biological environment, for example, kinesin walking on the microtubule[61] or during an experimental pulling/deformation/indentation measurement.

The running of Langevin simulations with implicit approxi-mate or exact treatments of hydrodynamic interactions on a GPU is now possible with the SOP-GPU package fully available for downloading at https://github.com/BarsegovGroup/SOP-GPU.

Acknowledgments

[1] T. Frembgen-Kesner, A. H. Elcock, *J. Chem. Theory Comput.* **2009**, *5*, 242.

[2] U. Winter, T. Geyer, *J. Chem. Phys.* **2009**, *131*, 104102.

[3] D. Brune, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2930.

[4] T. Frembgen-Kesner, A. H. Elcock, *Biophys. J.* **2010**, *99*, L75.

[5] M. Cieplak, S. Niewieczerzal, *J. Chem. Phys.* **2009**, *130*, 124906.

[6] P. Szymczak, M. Cieplak, *J. Phys. Condens. Matter.* **2007**, *19*, 285224.

[7] C. Hyeon, R. I. Dima, D. Thirumalai, *Structure* **2006**, *14*, 1633.

[8] M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, M. Rief, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20268.

[9] J. C. Lin, D. Thirumalai, *J. Am. Chem. Soc.* **2008**, *130*, 14080.

[10] C. Hyeon, P. A. Jennings, J. A. Adams, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3023.

[11] R. Tehver, D. Thirumalai, *Structure* **2010**, *18*, 471.

[12] A. Zhmurov, R. I. Dima, Y. Kholodov, V. Barsegov, *Proteins* **2010**, *78*, 2984.

[13] A. Zhmurov, K. Rybnikov, Y. Kholodov, V. Barsegov, *J. Phys. Chem. B* **2011**, *115*, 5278.

[14] A. Zhmurov, O. Kononova, R. I. Litvinov, R. I. Dima, V. Barsegov, J. W. Weisel, *J. Am. Chem. Soc.* **2012**, *134*, 20396.

[15] A. Zhmurov, A. E. X. Brown, R. I. Litvinov, R. I. Dima, J. W. Weisel, V. Barsegov, *Structure* **2011**, *19*, 1615.

[16] O. Kononova, Y. Kholodov, K. E. Theisen, K. A. Marx, R. I. Dima, F. I. Ataullakhanov, E. L. Grishchuk, V. Barsegov, *J. Am. Chem. Soc.* **2014**, *136*, 17036.

[17] O. Kononova, J. Snijder, M. Brasch, J. Cornelissen, R. I. Dima, K. A. Marx, G. J. Wuite, W. H. Roos, V. Barsegov, *Biophys. J.* **2013**, *105*, 1893.

[18] O. Kononova, J. Snijder, Y. Kholodov, K. A. Marx, G. J. Wuite, W. H. Roos, V. Barsegov, *PLoS Comput. Biol.* **2016**, *12*, e1004729.

[19] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, K. Schulten, *J. Comput. Chem.* **2007**, *28*, 2618.

[20] J. Anderson, C. Lorenz, A. Travesset, *J. Comput. Phys.* **2008**, *227*, 5342.

[21] M. J. Harvey, G. Giupponi, G. De Fabritiis, *J. Chem. Theory Comput.* **2009**, *5*, 1632.

[22] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2012**, *8*, 1542.

[23] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, *9*, 3878.

[24] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *12*, 19.

[25] D. L. Ermak, J. A. McCammon, *J. Chem. Phys.* **1978**, *69*, 1352.

[26] M. Fixman, *Macromolecules* **1981**, *14*, 1710.

[27] T. Ando, E. Chow, Y. Saad, J. Skolnick, *J. Chem. Phys.* **2012**, *137*, 064106.

[28] T. Geyer, U. Winter, *J. Chem. Phys.* **2009**, *130*, 114905.

[29] C. W. Oseen, Hydrodynamik; Akad. Verl.-Ges., Leipzig, **1927**, p 337.

[30] J. Rotne, S. Prager, *J. Chem. Phys.* **1969**, *50*, 4831.

[31] H. Yamakawa, *J. Chem. Phys.* **1970**, *53*, 436.

[32] Z. Liang, Z. Gimbutas, L. Greengard, J. Huang, S. Jiang, *J. Comput. Phys.* **2013**, *234*, 133.

[33] E. Smith, I. Snook, W. van Megen, *Phys. A* **1987**, *143*, 441.

[34] R. R. Schmidt, J. G. H. Cifre, J. G. de la Torre, *J. Chem. Phys.* **2011**, *135*, 084116.

[35] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, GNU Scientific Library Reference Manual, 3rd ed.; Net-work Theory Ltd., Surrey, United Kingdom, **2009**, p 592.

[36] J. M. Kollman, L. Pandi, M. R. Sawaya, M. Riley, R. F. Doolittle, *Biochem-istry* **2009**, *48*, 3877.

[37] R. Ranganathan, K. P. Lu, T. Hunter, J. P. Noel, *Cell* **1997**, *89*, 875.

[38] G. Spraggon, S. J. Everse, R. F. Doolittle, *Nature* **1997**, *389*, 455.

[39] E. Nogales, M. Whittaker, R. A. Milligan, K. H. Downing, *Cell* **1999**, *96*, 79.

[40] N. W. Park, S. R. Boduluri, J. F. Moomaw, P. J. Casey, L. S. Beese, *Science* **1997**, *275*, 1800.

[41] M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban, *Nat. Struct. Mol. Biol.* **2008**, *15*, 939.

[42] T. Veitshans, D. Klimov, D. Thirumalai, *Fold. Des.* **1997**, *2*, 1.

[43] J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932.

[44] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson, G. De Fabritiis, *J. Chem. Inf. Model.* **2010**, *50*, 397.

[45] O. Kononova, R. I. Litvinov, A. Zhmurov, A. Alekseenko, C. H. Cheng, S. Agarwal, K. A. Marx, J. W. Weisel, V. Barsegov, *J. Biol. Chem.* **2013**, *288*, 22681.

[46] A. Kazennov, A. Alekseenko, V. Bozhenko, T. Kulinich, N. Shuvalov, Y. Kholodov, *Comput. Mol. Biosci.* **2013**, *03*, 53.

[47] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 1011.

[48] A. Grossfield, WHAM: The weighted histogram analysis method, version 2.0.7. http://membrane.urmc.rochester.edu/content/wham (accessed Dec 1, 2014)

[49] M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly, M. Gruebele, *J. Mol. Biol.* **2001**, *311*, 373.

[50] E. Bura, D. K. Klimov, V. Barsegov, *Biophys. J.* **2007**, *93*, 1100.

[51] A. Zhmurov, R. I. Dima, V. Barsegov, *Biophys. J.* **2010**, *99*, 1259.

[52] N. Ferguson, C. M. Johnson, M. Macias, H. Oschkinat, A. Fersht, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13002.

[53] H. Nguyen, M. Jäger, A. Moretto, M. Gruebele, J. W. Kelly, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3948.

[54] O. Kononova, L. Jones, V. Barsegov, *J. Chem. Phys.* **2013**, *139*, 121913.

[55] O. Kononova, A. Zhmurov, K. A. Marx, V. Barsegov, In *Handbook Coarse-grained Modeling of Biomolecules*; G. A. Papoian, Ed.; Taylor & Francis, **2016**; Chapter 17 (in press).

[56] P. I. Zhuravlev, G. Reddy, J. E. Straub, D. Thirumalai, *J. Mol. Biol.* **2014**, *426*, 2653.

[57] J. C. Lin, D. Thirumalai, *J. Am. Chem. Soc.* **2013**, *135*, 16641.

[58] M. Hinczewski, J. C. M. Gebhardt, M. Rief, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4500.

[59] W. Lee, X. C. Zeng, K. Rotolo, M. Yang, C. J. Schofield, V. Bennett, W. T. Yang, P. E. Marszalek, *Biolphys. J.* **2012**, *102*, 1118.

[60] J. Y. Lee, T. M. Iverson, R. I. Dima, *J. Phys. Chem. B* **2011**, *115*, 186.

[61] Y. Goldtzvik, Z. Zhang, D. Thirumalai, *J. Phys. Chem. B* **2016**, 120, 2071.

[62] G. Hummer, *New J. Phys.* **2005**, *7*, 34.